MAPer: A **M**ulti-scale **A**daptive **Per**sonalized Model for Temporal Human Behavior Prediction

Sarah Masud Preum

John A. Stankovic

Yanjun Qi

UNIVERSITY of VIRGINIA

Modeling Temporal Human Behavior

- Generation of human behavior data
  - Online behavior:
    - social media, search log
    - Targeted advertising/ content sharing
    - Personalized IR
  - Offline behavior:
    - Sensors, smart devices
    - **Predicting** occupancy and energy usage
    - Anomaly detection in assisted living facilities

Factors Affecting Regular Temporal Behavior

- Temporal smoothness (lag)
  - Working from 3 pm to 4pm
    (and then continue after 4pm)
- Behavior rhythm (cycle)
  - Watching TV at every Saturday night
- Interaction among multiple activities
  - Working till late night delays sleep time

Fa
inc

Pe
m

Dynamic Nature Of Behavior

– Factors vary over multi-scale temporal contexts
  - Hour of the day
  - Day of the week

Adaptive modeling

Contribution: **M**ulti-scale **A**daptive **Per**sonalized Model (**MAPer**)

- Extracts features from major temporal factors
- Encodes multi-scale temporal contexts to ensure adaptive learning
- A linear predictive model with explanatory power

Solution Overview

Behavior time series

↓

Behavior Matrix

↓

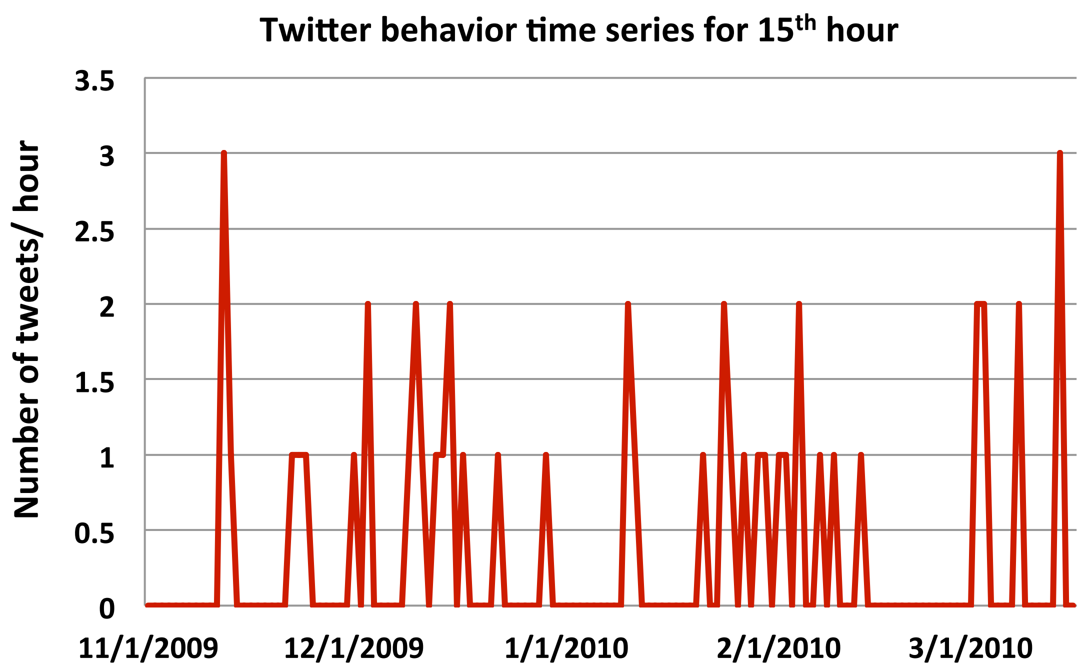
Extract features: **lag, cycle, interaction, temporal context**

↓

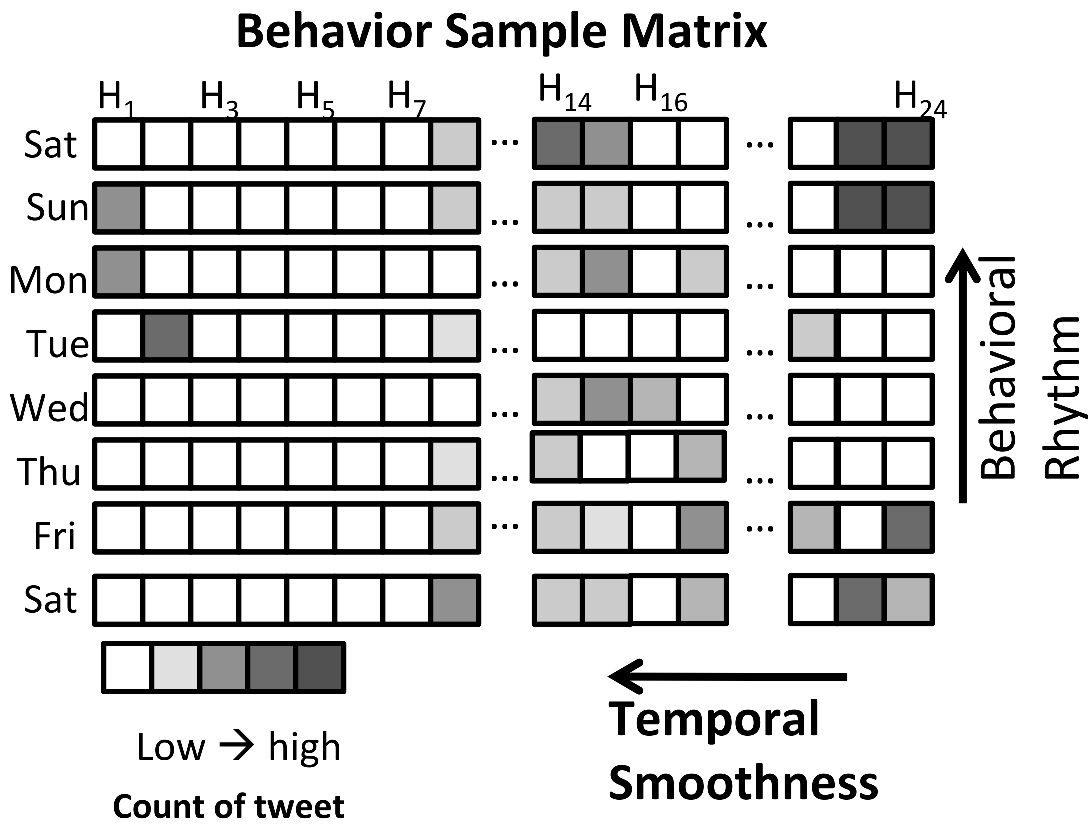
Use features for prediction

Creating Behavior Time Series

- Quantify behavior in the temporal domain as discrete behavior sample



Creating Behavior Sample Matrix



Lag and Cycle Features

- Lag of order i at time y_t : y_{t-i}

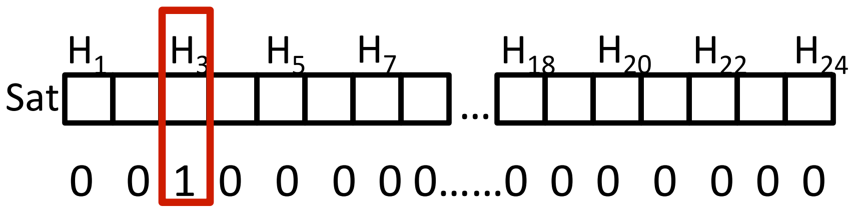
$$\hat{y}_t \approx \sum_{i=1}^{l} \alpha_i \cdot y_{(t-i)}$$

- Cycle of behavior time series

$$\hat{y}_t \approx \sum_{i=1}^{l} \alpha_i \cdot y_{(t-i)} + \boxed{\sum_{j=1}^{c} \beta_j \cdot y_{(t-cj)}}$$

Temporal Context Features

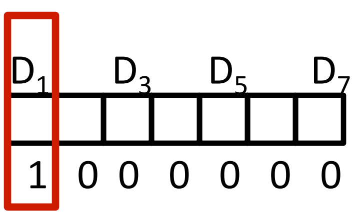
- Daily basis vector: $\vec{B}_d$



Example: Basis vector for hour 3

- Weekly basis vector: $\vec{B}_w$

Quantify the effect of temporal context on behavior



Example: Basis vector for Saturday

Features for a Single Activity

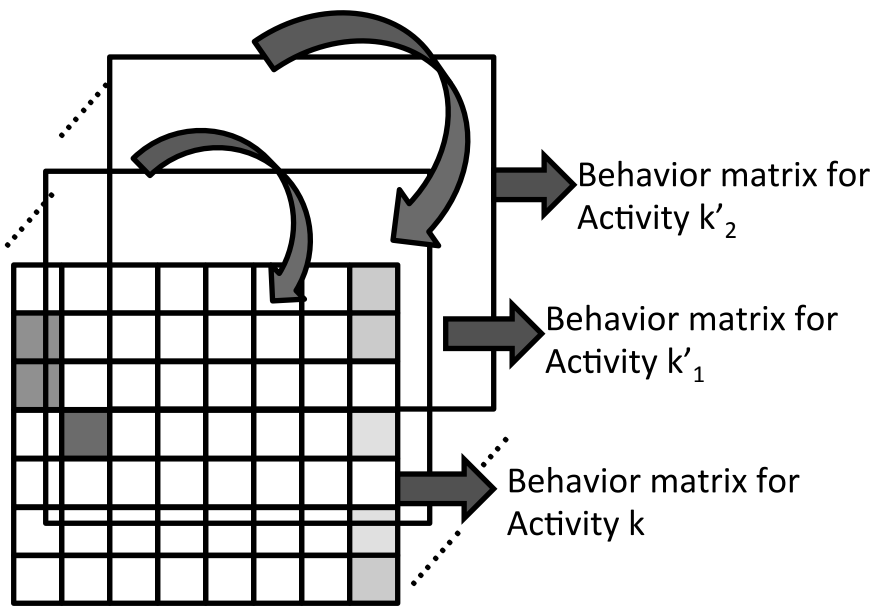
$$\underbrace{\hat{y}_t^k \approx \sum_{i=1}^{L_k} \alpha_i^k \cdot \hat{y}_{t-i}^k}_{\text{Lag}} + \underbrace{\sum_{j=1}^{C_k} \beta_j^k \cdot \hat{y}_{t-c_j}^k}_{\text{Cycle}} + \underbrace{\gamma^k \cdot (\vec{\mathrm{B}}_d, \vec{\mathrm{B}}_w)}_{\text{Temporal Context}}$$

Only daily basis vector: $\vec{B}_d$
Daily scale Adaptive Personalized model (**DAP**er)

Only weekly basis vector: $\vec{B}_w$
Weekly scale Adaptive Personalized model (**WAP**er)

Both $\vec{B}_d$ and $\vec{B}_w$:
Multi-scale Adaptive Personalized model (**MAP**er)

Interaction Features for Multiple Activities



activity	interactions
R1 sleep	Brushing teeth Using internet Watching TV
R2 sleep	Brushing teeth Watching TV Washing dishes
R1 snack	Watching TV Using internet Talking on phone
…	…

$$\hat{y}_t^k \approx \sum_{i=1}^{L_k} \alpha_i^k \cdot y_{t-i}^k + \sum_{j=1}^{C_k} \beta_j^k \cdot y_{t-c_j}^k + \gamma^k(\vec{\mathrm{B}}_d, \vec{\mathrm{B}}_w) + \sum_{k' \in S} \sum_{l'=1}^{L_{k'}} W_{l'}^{(k,k')} \cdot y_{t-l'}^{k'}$$

interaction

Related Works

- Time series prediction
  - Seasonal ARIMA (SARIMA) model
  - Does not consider temporal context of behavior
- Modeling online user behavior
  - Search query [K. Radnisky et al, WWW 2012; J. Yang et al, WSDM 2011]
  - Social media posts [F. Abel et al, UMAP 2011]
  - Focus on temporal pattern of user generated contents rather than actual user behavior
- Modeling offline user behavior
  - Computer vision and WSN: Sensing and recognizing different activities of daily living
  - Don't focus on predicting

Overview of Evaluation

- Predicting behavior intensity at a time interval as a regression problem
  - Performance metric: MSE, Pearson Correlation
- 4 real datasets
- Comparison with
  - Parametric and non parametric baselines
  - state-of-art SARIMA model
- Sensitivity analysis

Datasets

Online Behavior Data

Dataset	Span	# of users	Behavior sample
Twitter	5 months	**1274**	# of tweets /hour
Search log	3 months	**1307**	# of unique search queries /hour

Offline Behavior Data

Dataset	Span	# of resident	Behavior sample
ARAS	1 month	**2**	Duration of activity/half hour
HOLMES	3 months	**1**	

Baselines

- Moving average over both lag and cycle terms (**MA**)

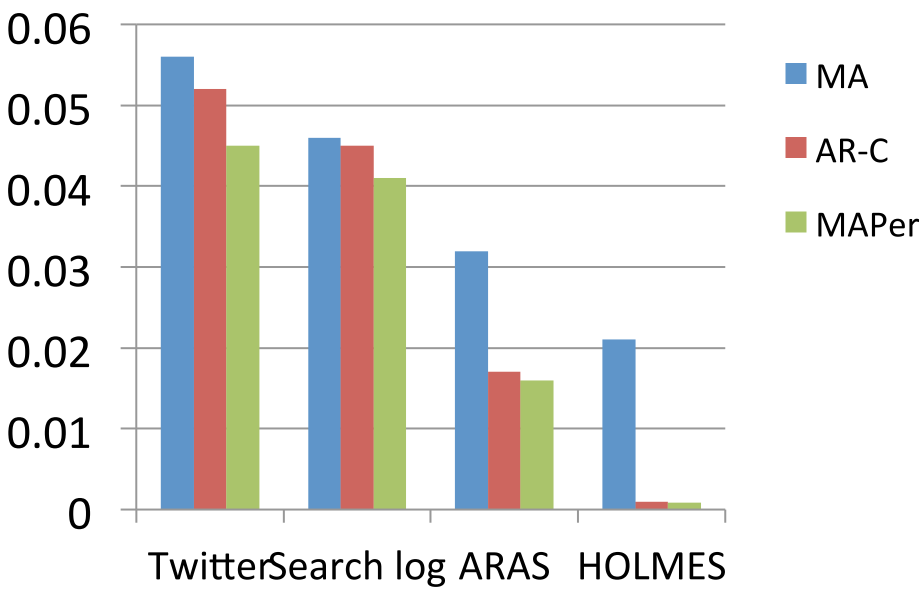
$$\hat{y}_t = \frac{1}{2}\left(\frac{\sum_{i=1}^{l} y_{t-i}}{l} + \frac{\sum_{j=1}^{c} y_{(t-j*24)}}{c} \right)$$

- Auto-regressive method with cycle feature (**AR-C**)

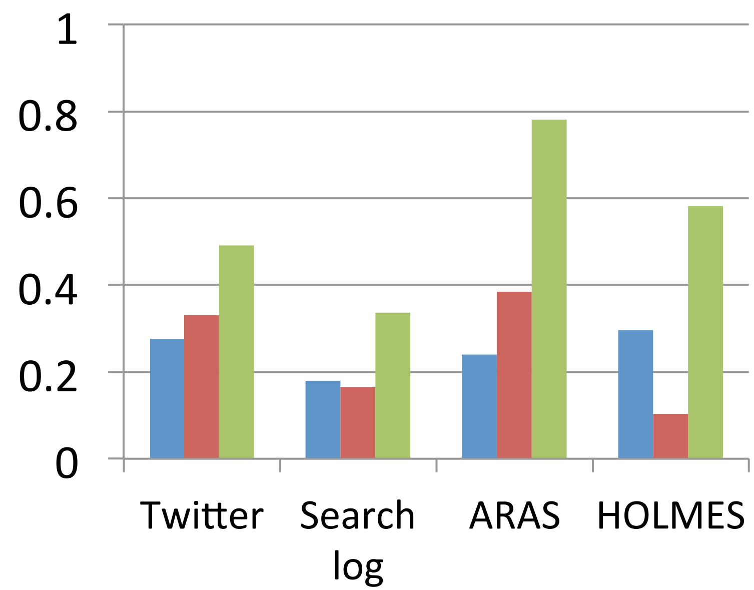
$$\hat{y}_t \approx \sum_{i=1}^{l} \alpha_i \cdot y_{(t-i)} + \sum_{j=1}^{c} \beta_j \cdot y_{(t-cj)}$$

Comparing with Baselines

- On average, MAPer reduces MSE by 10% and increases Pearson correlation by 83%



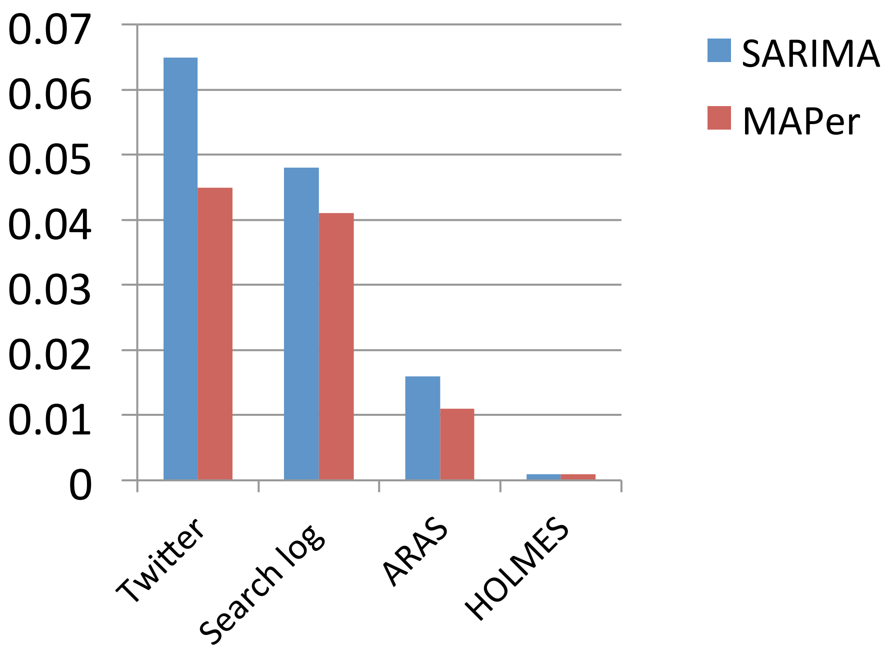
Mean Square Error



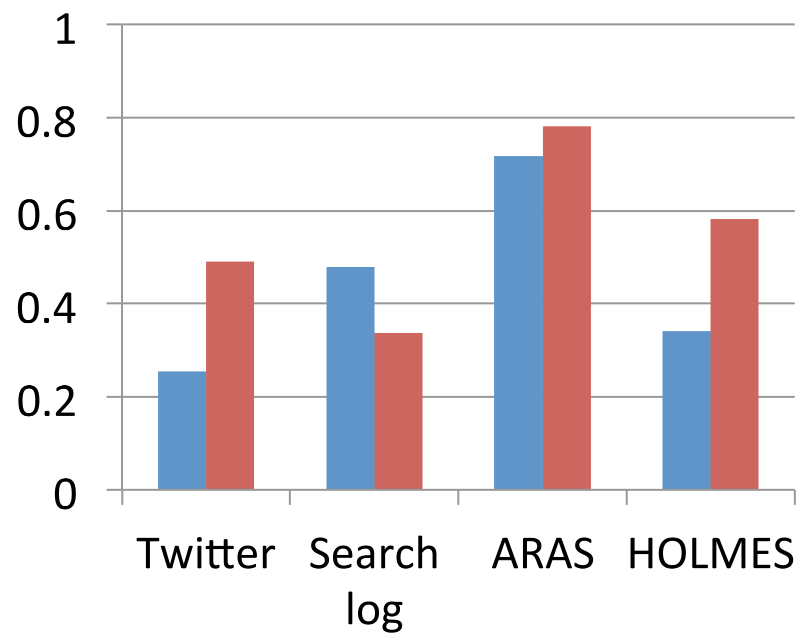
Pearson Correlation

Comparing with State of Art

- On average, MAPer reduces MSE by 14% and increases Pearson Correlation by 44% than SARIMA



Mean Square Error



Pearson Correlation

Effect of Temporal Context Features

- Daily context is more useful than weekly context

Dataset	Pearson Correlation			Mean Square Error		
	DAPer	WAPer	MAPer	DAPer	WAPer	MAPer
Search log	0.33	0.21	**0.34**	0.041	0.044	**0.041**
Twitter	0.49	0.40	**0.49**	0.045	0.048	**0.045**
ARAS	**0.78**	0.68	0.78	0.016	0.019	0.016
HOLMES	**0.58**	0.33	0.58	0.0009	0.0009	0.0009

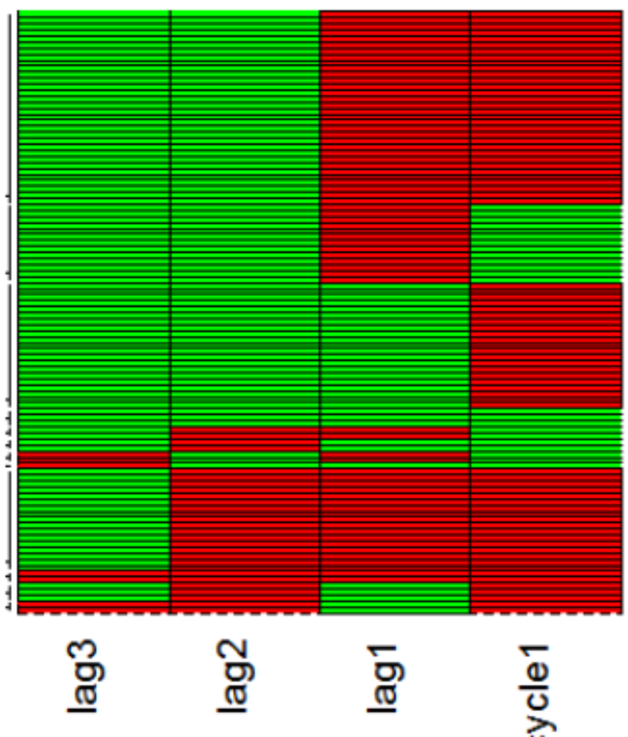
Effect of Interaction Features

- Interaction features improves the performance

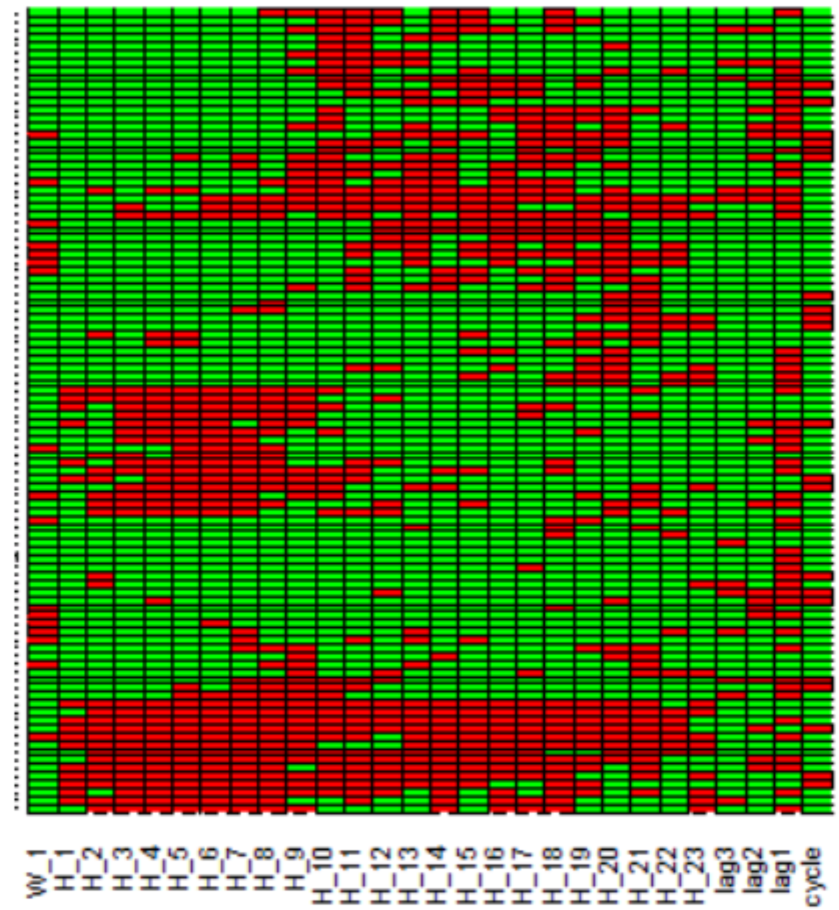
	Pearson Correlation		Mean Square Error	
	ARAS	HOLMES	ARAS	HOLMES
MAPer w/o Interaction	0.7169	0.385	0.0184	0.0009
MAPer	0.782	0.583	0.016	0.0009
%Improvement	9	51	15	-

Results: Explanatory Power of MAPer

- Quantify effect of different features
- Detecting user similarity more precisely



(a) AR-C model



(b) MAPer model

Insights from results

Experiment	Online Behavior	Offline Behavior
Prediction	MAPer	MAPer with interation
Personalization	✓	NA
Adaptive Learning	✓	✓
Variation of temporal window length	Higher better	Varies for each activity: *having snack vs sleep*
Variation of training set size	Lower better	
Variation of lag	No significant effect	

Concluding Remarks

- A personalized interpretable model for temporal behavior prediction
- Virtual and physical behavior
- Some regularity in behavior that conforms with hours of the day, day of the week

Thanks!

Backup

Dataset Deatails

Dataset	Total users considered	Span	Minimum days used	Filtered users	Format	Behavior sample
Twitter, CIKM 2009	120,689	5 months	>=120 days	**1274**	USER_ID, Text, Timestamp	# of tweets / hour
Microsoft AOL Search log	182,348	3 months	>=60 days	**1307**	USER_ID, Query, QueryTime, ItemRank, ClickURL	#of unique search queries /hour

Datasets: Activities of daily living (ADL)

- ARAS Smart home Data
  - 2 person resident with **27** labeled activities per person spanning over **1 month**
  - Sleeping, having breakfast/dinner/lunch, using internet, watching TV, reading, using phone, toileting
  - Format: {Day, activity name, start time, end time, resident_ID}
  - Duration of an activity per half hour as behavior sample

- Predict health-relevant activities
  - Sleeping, having snack, breakfast, lunch, dinner

Comparing with State of Art

- On average, MAPer reduces MSE by 14% and increases Pearson Correlation by 44% than SARIMA

Dataset	Baseline Methods		state-of-art		Multi-scale Adaptive Methods		% Improvement over state-of-art
	MA	AR-C	Global SARIMA	Local SARIMA	Global MAPer	Local MAPer	
Search log	0.046	0.045	0.065	0.048	0.045	**0.041**	14.6
Twitter	0.056	0.052	0.073	0.065	0.049	**0.045**	30.7
ARAS	0.032	0.017	0.042	**0.011**	0.018	0.016	-
HOLMES	0.021	0.001	0.082	0.001	0.001	**0.0009**	10

Dataset	Baseline Methods		state-of-art		Multi-scale Adaptive Methods		% Improvement over state-of-art
	MA	AR-C	Global SARIMA	Local SARIMA	Global MAPer	Local MAPer	
Search log	0.180	0.165	0.098	**0.479**	0.228	0.337	-
Twitter	0.277	0.330	0.079	0.254	0.410	**0.491**	93.3
ARAS	0.239	0.384	0.204	0.7166	0.446	**0.782**	9
HOLMES	0.296	0.104	0.136	0.340	0.277	**0.583**	71.5

Effect of Personalization

- Personalization improves the performance drastically

	Pearson Correlation		Mean Square Error	
	Search log	Twitter	Search log	Twitter
Personalized	**0.337**	**0.491**	**0.041**	**0.045**
Aggregated	0.127	0.192	0.114	0.064
Improvement	2.65 times	2.56 times	64%	30%

Table 5: Personalized models for each user significantly improve the prediction performance for both Twitter and search log datasets.

Sensitivity Analysis

Parameter	Range	
	Online Behavior	Offline Behavior
Scale (hours)	**1**,2,4,6	**1/2**,1,2,4
Training set size (weeks)	2,**4**,6,8	NA[6]
Lag (hours)	2,3,**4**,6,8	1/2,1,**2**,3,4

Table 8: The range of different parameters used in experiments. The default values of parameters are shown in bold.

Effect of Varying Training Set Size

- The most recent behavior data is more important

Methods	Pearson Correlation		Mean Square Error	
	Search log	**Twitter**	**Search log**	**Twitter**
2 weeks	**0.4**	**0.52**	**0.039**	**0.044**
4 weeks	0.34	0.49	0.041	0.045
6 weeks	0.31	0.45	0.042	0.048
8 weeks	-	0.44	-	0.048
% Improvement	31	18	7	8

Table 9: Training set with more recent data results into better prediction.

Effect of Varying Temporal Window

Scale	Pearson Correlation		Mean Square Error	
	Search log	**Twitter**	**Search log**	**Twitter**
1 hour	0.337	0.49	0.041	0.045
2 hours	0.38	0.45	0.022	0.0261
4 hours	0.427	0.49	0.011	0.0138
6 hours	**0.431**	**0.50**	**0.008**	**0.0094**
Improvement	27%	11%	5 times	4.7 times

Table 10: Varying temporal scale: larger scale results into better prediction performance for online behavior data.

Interactive Activities

Activity	Influential Activities
R1:sleep	brushing teeth, using internet, watching TV
R2:sleep	brushing teeth, watching TV, washing dishes
R1:snack	watching TV, using internet, talking on the phone
R2:snack	watching TV, having shower, using internet
R1:breakfast	preparing breakfast, using internet, watching TV
R1:lunch	preparing lunch, talking on the phone, watching TV
R1:dinner	talking on the phone, watching TV, preparing dinner

Table 6: Interaction among different activities of ARAS dataset: the left column contains the activities we want to predict and the right column contains the activities that interact with them.

Future work

- Forecast 1 step → forecast n steps
- Formulate as a multi-class classification problem
- Application level: Aggregating individual level models over a population/geographical region for predicting peak in web traffic

Distribution of Cycles

Dataset	Most frequent	2nd most freq	3 rd most freq	others
Search log	7 Days (627)	14 Days (171)	2 hours (15) 1 hour (12)	1,6,8 days
Twitter	7 Days (575)	1 hour (181)	14 days (74)	2,3,20 hours

ADL dataset insights

Activity	Best method	Best scale	Potential reason
R1_sleep	I-MAPer	½ hour	Regular timing, all most all methods perform well
R2_sleep	I-MAPer	½ hour	
R1_snack	AR-C	2 hour	No temporal context, prediction performance low
R1_breakfast	I-MAPer	2 hour	
R1_lunch	AR-C	2 hour	No temporal context
R1_dinner	I-MAPer	2 hour	
R2_snack	I-MAPer	2 hour	

Variation of scale: individual activity from ADL data in terms of PC

- Sleep pattern of 1 quite regular (around 2pm-10am)
- Meal time varies, so relaxing the scale helps

	½ hr	1 hr	2 hr	4 hr
R1_snack	-0.011	0.118	**0.181**	0.127
R1_sleep	**0.939**	0.910	0.904	0.902
R1_breakfast	0.495	0.223	**0.675**	0.631
R1_lunch	0.272	0.224	**0.435**	0.388
R1_dinner	0.228	0.102	**0.503**	0.490
R2_snack	0.292	0.333	0.507	**0.581**
R2_sleep	**0.883**	0.826	0.777	0.737

Variation of scale: individual activity from ADL data

- MSE

	½ hr	1 hr	2 hr	4 hr
R1_snack	0.023	0.014	0.007	0.003
R1_sleep	0.025	0.043	0.037	0.029
R1_breakfast	0.011	0.015	0.004	0.002
R1_lunch	0.007	0.004	0.002	0.001
R1_dinner	0.007	0.005	0.001	0.001
R2_snack	0.017	0.009	0.005	0.002
R2_sleep	0.030	0.050	0.057	0.059

ARIMA model

- Forecast at time t =

constant+

weighted sum of last p values of y+ **AR Terms**

weighted sum of last q forecast errors **MA Terms**

- ARIMA(p,d,q)

$$y_t = \mu + (\phi_1 y_{(t-1)} + \phi_2 y_{(t-2)} + \ldots + \phi_p y_{(t-p)})$$

$$+ (\theta_1 e_{(t-1)} + \theta_2 e_{(t-2)} + \ldots + \theta_q e_{(t-q)})$$

ARIMA model

- ACF that dies out gradually and PACF that cuts off sharply after a few lags →AR signature
  - An AR series is usually positively autocorrelated at lag 1
- ACF that cuts off sharply after a few lags and PACF that dies out more gradually → MA signature
  - An MA series is usually negatively autcorrelated at lag 1 (or even mildly overdifferenced)

ARIMA model

- Our model is seasonal ARIMA model (p,d,q)x(P,D,Q): (l,0,0)X(c,0,0)
- l: lag, c:cycle
- d=0, D=0: as we assumed the time series to be stationary
  – we did not perform any differencing

ARIMA forecasting equation

- Let Y denote the *original* series
- Let y denote the *differenced* (stationarized) series

o difference (d=0): $y_t = Y_t$

rst difference (d=1): $y_t = Y_t - Y_{t-1}$

econd difference (d=2): $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$

$$= Y_t - 2Y_{t-1} + Y_{t-2}$$

lote that the second difference is not just the change relative to two eriods ago, i.e., it is *not* $Y_t - Y_{t-2}$. Rather, it is the change-in-the-change, hich is a measure of local "acceleration" rather than trend.

Undifferencing the forecast

The differencing (if any) must be *reversed* to obtain a forecast for the original series:

$$\text{If } d = 0: \quad \hat{Y}_t = \hat{y}_t$$

$$\text{If } d = 1: \quad \hat{Y}_t = \hat{y}_t + Y_{t-1}$$

$$\text{If } d = 2: \quad \hat{Y}_t = \hat{y}_t + 2Y_{t-1} - Y_{t-2}$$

Fortunately, your software will do all of this automatically!